

**Original citation:**

Duong, Manh Hong and Tran, Minh Hoang. (2017) Analysis of the mean squared derivative cost function. Mathematical Methods in the Applied Sciences.

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/86313>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

This is the peer reviewed version of the following article: Duong, M. H., and Tran, H. M. (2017) Analysis of the mean squared derivative cost function. Math. Meth. Appl. Sci., doi: 10.1002/mma.4382., which has been published in final form at <https://doi.org/10.1002/mma.4382>. This article may be used for non-commercial purposes in accordance with [Wiley Terms and Conditions for Self-Archiving](#).

**A note on versions:**

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

# Analysis of the mean squared derivative cost function

Manh Hong Duong<sup>1</sup> and Minh Hoang Tran<sup>2,3</sup>

<sup>1</sup>Mathematics Institutes, University of Warwick, Coventry CV4 7AL.

Email: m.h.duong@warwick.ac.uk.

<sup>2</sup>Department of Industrial Engineering, Texas A&M University, College Station, TX 77843. Email: tran@tamu.edu

<sup>3</sup>School of Applied Mathematics and Informatics, Hanoi University of Science & Technology, Hanoi, Vietnam

February 28, 2017

## Abstract

In this paper, we investigate the mean squared derivative cost functions that arise in various applications such as in motor control, biometrics and optimal transport theory. We provide qualitative properties, explicit analytical formulas and computational algorithms for the cost functions. We also perform numerical simulations to illustrate the analytical results. In addition, as a by-product of our analysis, we obtain an explicit formula for the inverse of a Wronskian matrix that is of independent interest in linear algebra and differential equations theory.

## 1 Introduction

### 1.1 The mean squared derivative cost function

This paper is concerned with the analysis of *the mean squared derivative cost function* defined as follows:

$$C_{n,h}(x_0, x_1, \dots, x_{n-1}; y_0, y_1, \dots, y_{n-1}) := \inf_{\xi} \int_0^h |\xi^{(n)}(t)|^2 dt, \quad (1)$$

where  $\mathbf{x} = (x_0, \dots, x_{n-1}) \in \mathbb{R}^{dn}$ ,  $\mathbf{y} = (y_0, \dots, y_{n-1}) \in \mathbb{R}^{dn}$ , the infimum is taken over all curves  $\xi \in C^n([0, h], \mathbb{R}^d)$  that satisfy the boundary conditions

$$(\xi, \xi', \dots, \xi^{(n-1)})(0) = (x_0, x_1, \dots, x_{n-1}) \quad \text{and} \quad (\xi, \xi', \dots, \xi^{(n-1)})(h) = (y_0, y_1, \dots, y_{n-1}). \quad (2)$$

Throughout this paper,  $h$  is a constant representing the final time.

This cost function plays a central role in various practical applications and theoretical research. For the motivation of this paper, we now review some important literature here.

## 1.2 Literature review and motivation of the present paper

*Applications in engineering and applied sciences.* In the literature, the minimization problem (1) for  $n = 2, 3, 4, 5, 6$  is respectively called the principle of minimum acceleration, jerk, snap, crackle, and pop; see for instance [25]. The minimal jerk principle was initially used to model velocity profiles generated by elbow movements [20] and later extended to trajectory prediction for reaching movements between visual targets in the horizontal plane and to curved and obstacle-avoidance movements [16]. Since then the minimal jerk principle and other mean squared derivative cost functions (1) have been found to be useful in the modelling and design of various real-world systems. Examples of such applications include motor control [15], biometrics and online-signatures [4, 5] and robotics [17], just to name a few. We refer to the mentioned papers and references therein for more information. Since the cost function has applications in many different contexts, a thorough analysis and computational method for the general case, i.e. with arbitrary  $n$  and boundary values  $\mathbf{x}$  and  $\mathbf{y}$ , would be useful. For instance, it can be pre-computed and embedded in a larger algorithm. However, in the literature there exist neither analytic formulas nor computational methods for the general case.

**Question 1.** *Can one establish an explicit formula and an algorithm to compute the cost function  $C_{n,h}$  for the general case?*

*Usage in theoretical research.* The mean squared derivative cost functions also appear in theoretical research such as in partial differential equations theory and optimal transport theory. When  $n = 1$ , the function

$$\Phi_1(t, x, y) = \frac{1}{(4\pi t)^{d/2}} \exp\left(-\frac{C_{1,t}(x, y)}{4}\right)$$

is the fundamental solution of the diffusion/heat equation  $\partial_t u = \Delta u$ . When  $n = 2$ , the function

$$\Phi_2(t, x, x'; y, y') = \frac{\beta_d}{t^{2d}} \exp\left(-\frac{C_{2,t}(x, x'; y, y')}{4}\right),$$

where  $\beta_d$  is a normalising constant, is the fundamental solution of the following ultra-parabolic equation

$$\partial_t u + y \cdot \nabla_x u = \Delta_y u. \quad (3)$$

where subscripts in  $\nabla_x$  and  $\Delta_y$  indicate that these differential operators act only on those variables. This equation was first studied by Kolmogorov [24] and then was Hörmander's starting point to develop the hypo-elliptic theory [21].

The diffusion/heat equation and the above ultra-parabolic equation are special cases of the following hypo-elliptic equation

$$\partial_t u + \sum_{i=1}^{n-1} x_{i+1} \cdot \nabla_{x_i} u = \Delta_{x_n} u. \quad (4)$$

Equations of this type have been studied from various points of view such as trends to equilibrium [14], Gaussian-estimates for the fundamental solution [8], connections to particle systems and coarse-graining [10, 12, 7].

A natural question arises

**Question 2.** *Is*

$$\Phi_n(t, \mathbf{x}, \mathbf{y}) := \frac{\tilde{\beta}_d}{t^{\frac{n^2 d}{2}}} \exp\left(-\frac{C_{n,t}(\mathbf{x}, \mathbf{y})}{4}\right), \quad \text{for some normalising constant } \tilde{\beta}_d, \quad (5)$$

*the fundamental solution of (4)?*

In the optimal transport theory, the cost function  $C_{n,h}$  between two points  $\mathbf{x}$  and  $\mathbf{y}$  in the Euclidean space  $\mathbb{R}^{dn}$  can be used to define a Monge-Kantorovich optimal transport cost function between two probability measures  $\mu(d\mathbf{x})$  and  $\nu(d\mathbf{y})$  on  $\mathbb{R}^{dn}$  as follows

$$\mathcal{W}_n^2(\mu, \nu) := \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{\mathbb{R}^{dn} \times \mathbb{R}^{dn}} C_{n,h}(\mathbf{x}; \mathbf{y}) \gamma(d\mathbf{x} d\mathbf{y}), \quad (6)$$

where  $\Gamma(\mu, \nu)$  denotes the set of all probability measures on  $\mathbb{R}^{dn} \times \mathbb{R}^{dn}$  having  $\mu$  and  $\nu$  as first and second marginals. The Monge-Kantorovich cost function is the central object in optimal transport theory with many applications in other fields of mathematics and economics; see for instance [26, 27] for nice expositions on optimal transport theory and its applications. In particular, for  $n = 1$ ,  $C_{1,h}(x_0, y_0) = \frac{1}{h}|y_0 - x_0|^2$  and  $\mathcal{W}_1$  is the well-known Wasserstein distance. Wasserstein gradient flows, i.e., gradient flows of energy functionals with respect to the Wasserstein metric, form an important class of dissipative evolution equations, see for instance [23, 2]. When  $n = 2$ ,  $C_{2,h}(x_0, x_1; y_0, y_1) = \frac{1}{h} \left[ |y_1 - y_0|^2 + 12 \left| \frac{x_1 - x_0}{h} - \frac{y_1 + y_0}{2} \right|^2 \right]$  and  $\mathcal{W}_2$  is the minimal acceleration cost function. This cost function has been used to construct variational formulation for the Kramers equation (Equation (3) above with additional terms coming from external and frictional forces) showing that the Kramer equation is a (generalised) gradient flow of the Boltzmann entropy with respect to the Monge-Kantorovich transport cost  $\mathcal{W}_2$  [22, 11]. In addition,  $\mathcal{W}_2$  has also been used in constructing variational schemes for other evolution equations such as the system of isentropic Euler equations [18, 28] and the compressible Euler equations [6].

**Question 3.** *Is Equation (4) a (generalised) gradient flow of the Boltzmann entropy with respect to the Monge-Kantorovich transport cost  $\mathcal{W}_n$ ?*

We provide further discussions on the motivation of the present paper in Section 2.2.

### 1.3 The aim of the present paper

The aim of this paper is to address Question 1. We develop analytic and computational aspects. We show some qualitative properties of  $C_{n,h}$  in Theorem 1.1; we provide an explicit analytical formula for  $C_{n,h}$  in Theorem 1.2; and we present a computational method for  $C_{n,h}$  in Theorem 1.3 and in Algorithm 5.1.

Questions 2 and 3 will be answered in a companion paper [13] where, using analytical formulas in Theorem 1.2 and Theorem 1.3 of the present paper, we will prove that the function  $\Phi$  defined in (5) is the fundamental solution of Equation (4) and show that this equation is indeed a (generalised) gradient flow of the Boltzmann entropy with respect to the Monge-Kantorovich transport cost  $\mathcal{W}_n$  via a variational approximation scheme.

### 1.4 Main results of the present paper

We now describe our main results. The detailed statements will be given in the subsequent sections. Our first result concerns the qualitative behaviour of the cost function as function of  $h$  and  $n$ .

**Theorem 1.1.** *The cost function  $C_{n,h}$  is scalable with respect to  $h$  and monotonically increasing with respect to  $n$ .*

The full description and proof of this theorem are given in Theorem 2.1 in Section 2. Moreover, we also provide an interpretation of the cost function based on the theory of large deviations.

Our second theorem is an explicit analytical formula for the cost function.

**Theorem 1.2.** *The cost function  $C_{n,h}$  has an explicit formula given by*

$$C_{n,h}(x_0, \dots, x_{n-1}; y_0, \dots, y_{n-1}) = \mathbf{b}_n(h)^T B_n(h) [A_n(h)]^{-1} \mathbf{b}_n(h), \quad (7)$$

where the vector  $\mathbf{b}_n(h)$  and the two matrices  $B_n(h)$  and  $A_n(h)$  are given explicitly in (20), (24) and (19) respectively.

The proof of this theorem is given in Section 3.

The last theorem provides explicit formulas for the  $LU$  decomposition of  $A_n$  and for  $A_n^{-1}$ .

**Theorem 1.3.** *The matrix  $A_n(h)$  has an  $LU$ -decomposition as in (31)-(32). The inverses of the matrices  $L$  and  $U$  are given explicitly in (33)-(34).*

We prove this theorem in Section 4 (cf. Theorem 4.1). As we show there, the matrix  $A_n(h)$  is a Wronskian matrix, which plays an important role in linear algebra and differential equations; hence, this theorem is of independent interest.

## 1.5 Organisation of the paper

The rest of this paper is structured as follows. In Section 2, we study some qualitative properties of the cost function  $C_{n,h}$ . In Section 3 we provide an explicit formula for  $C_{n,h}$ . The  $LU$ -decomposition of the matrix  $A$  is presented in Section 4. In Section 5 we provide an algorithm to compute the cost function and compute the expressions obtained for small  $n$  explicitly.

## 2 Qualitative properties of the cost functions

In this section, we provide the full description and proof of Theorem 1.1 on qualitative properties of the cost function, which characterises its behaviour as functions of  $n$  and  $h$ .

### 2.1 Behaviour of $C_{n,h}$ as a function of $n$ and $h$

**Theorem 2.1** (Qualitative properties of the cost function).

1. (*Scaling property of the cost functions*). It holds that

$$C_{n,h}(x_0, x_1, \dots, x_{n-1}; y_0, y_1, \dots, y_{n-1}) = h^{1-2n} \inf_{\xi} \int_0^1 |\xi^{(n)}(t)|^2 dt, \quad (8)$$

where the infimum is taken over the curves  $\xi \in C^n([0, 1], \mathbb{R}^d)$  such that

$$(\xi, \xi', \dots, \xi^{(n-1)})(0) = (x_0, hx_1, \dots, h^{n-1}x_{n-1}), \quad (\xi, \xi', \dots, \xi^{(n-1)})(1) = (y_0, hy_1, \dots, h^{n-1}y_{n-1}).$$

2. (*Monotonicity of the cost function*). It holds that

$$C_{n-1,h}(x_1, \dots, x_{n-1}; y_1, \dots, y_{n-1}) \leq C_{n,h}(x_0, x_1, \dots, x_{n-1}; y_0, y_1, \dots, y_{n-1}). \quad (9)$$

Note the difference between the right-hand sides of (1) and (8). In the former, the dependence on  $h$  appears in the interval of the integral, while in the latter, the dependence is moved to the boundary conditions. The pre-factor is also rescaled accordingly.

*Proof.* We first prove the first part. The assertion is simply followed from the change of variables:  $t \mapsto \tilde{t} := \frac{t}{h}$ . Define  $\tilde{\xi}(\tilde{t}) := \xi(t) = \xi(h\tilde{t})$ . Then for any  $0 \leq k \leq n$  we have

$$\tilde{\xi}^{(k)}(\tilde{t}) = h^k \xi^{(k)}(t).$$

Substituting this into the integral, we obtain

$$\int_0^h |\xi^{(n)}(t)|^2 dt = h^{1-2n} \int_0^1 |\tilde{\xi}^{(n)}(\tilde{t})|^2 d\tilde{t},$$

and the boundary conditions become

$$\begin{aligned}(\tilde{\xi}, \tilde{\xi}', \dots, \tilde{\xi}^{(n-1)})(0) &= (\xi, h\xi', \dots, h^{n-1}\xi^{(n-1)})(0) = (x_0, hx_1, \dots, h^{n-1}x_{n-1}), \\(\tilde{\xi}, \tilde{\xi}', \dots, \tilde{\xi}^{(n-1)})(1) &= (\xi, h\xi', \dots, h^{n-1}\xi^{(n-1)})(h) = (y_0, hy_1, \dots, h^{n-1}y_{n-1}).\end{aligned}$$

The assertion (8) then follows from these computations.

Next we prove the second statement (9). The minimizing problem (1) is of the form

$$\inf_{\xi} \int_0^h L(t, \xi, \xi', \dots, \xi^{(n)}) dt,$$

where  $L : [0, h] \times \mathbb{R}^{d(n+1)} \rightarrow \mathbb{R}$ ,  $L(t, x, p_1, \dots, p_n) = |p_n|^2$ . Since  $L$  depends only on  $p_n$  and  $p_n \mapsto L(t, x, p_1, \dots, p_n)$  is positive, continuous and convex; the existence and uniqueness of a minimizer follows from the direct method in the calculus of variations. Let  $\xi_{opt} \in C^n([0, h], \mathbb{R}^d)$  be the optimal curve in the definition of  $C_{n,h}(x_0, x_1, \dots, x_n; y_0, y_1, \dots, y_{n-1})$ . We define  $\eta(t) := \xi'_{opt}$ . Since  $\eta \in C^{n-1}([0, h], \mathbb{R}^d)$  and

$$(\eta, \dots, \eta^{(n-1)})(0) = (x_1, \dots, x_{n-1}), \quad (\eta, \dots, \eta^{(n-1)})(h) = (y_1, \dots, y_{n-1}),$$

it follows that  $\eta$  is an admissible curve in the definition of  $C_{n-1,h}(x_1, \dots, x_{n-1}; y_1, \dots, y_{n-1})$ . It implies that

$$\begin{aligned}C_{n-1,h}(x_1, \dots, x_{n-1}; y_1, \dots, y_{n-1}) &\leq \int_0^h |\eta^{(n-1)}(t)|^2 dt \\&= \int_0^h |\xi_{opt}^{(n)}(t)|^2 dt \\&= C_{n,h}(x_0, x_1, \dots, x_{n-1}; y_0, y_1, \dots, y_{n-1}).\end{aligned}$$

This finishes the proof of the theorem.  $\square$

The next lemma shows that among all the cost functions, only  $\sqrt{C_{1,h}}$  is a distance in the Euclidean space  $\mathbb{R}^{dn}$ .

**Lemma 2.2.**  $C_{n,h}(x_0, \dots, x_{n-1}; y_0, \dots, y_{n-1})$  is always non-negative and it equals 0 if and only if

$$y_j = \sum_{i=j}^{n-1} \frac{h^{i-j}}{(i-j)!} x_i \tag{10}$$

for  $j = 0, \dots, n-1$ . In particular, for any  $z \in \mathbb{R}^{dn}$  we have  $C_{n,h}(0, z) \geq 0$  and  $C_{n,h}(0, z) = 0$  iff  $z = 0$ .

*Proof.* This Lemma is a direct consequence of the definition of  $C_{n,h}$ . Obviously  $C_{n,h} \geq 0$  and it is equal to 0 if and only if the optimal curve satisfies

$$\xi^{(n)}(t) = 0, \tag{11}$$

i.e., it is a polynomial of order  $n-1$ . This together with the boundary conditions imply (10).  $\square$

**Remark 2.3.** Lemma 2.2 shows that  $\sqrt{C_{n,h}} : \mathbb{R}^{dn} \times \mathbb{R}^{dn} \rightarrow [0, \infty)$  is not a distance when  $n \geq 2$ . It is not symmetric and does not satisfy the condition  $\sqrt{C_{n,h}}(x_0, \dots, x_{n-1}; y_0, \dots, y_{n-1}) = 0$  iff  $(x_0, \dots, x_{n-1}) = (y_0, \dots, y_{n-1})$ .

## 2.2 Interpretation of the cost function based on large-deviation principles and further discussions

In this section, we provide an interpretation of the cost function based on a small-noise large-deviation principle for a special system of stochastic differential equations (SDEs) and further discussion. We consider a system of  $n$  coupled oscillators, each of them moving vertically and being connected to their nearest neighbours, the last oscillator being forced by a random noise. Mathematically, the system is given by the following system of SDEs with  $t \in [0, h]$ ,

$$\begin{aligned} d\xi &= \xi_2 dt \\ d\xi_2 &= \xi_3 dt \\ &\vdots \\ d\xi_{n-1} &= \xi_n dt \\ d\xi_n &= \sqrt{\varepsilon} dW(t), \end{aligned} \tag{12}$$

where  $W(t)$  is a  $d$ -dimensional Wiener process. The parameter  $\varepsilon$  represents the amplitude of the noise. This system can be formally written as

$$\xi_\varepsilon^{(n)}(t) = \sqrt{\varepsilon} \frac{dW}{dt}(t), \tag{13}$$

where the subscript indicates the dependence on  $\varepsilon$ . Suppose further that initial and terminal points are imposed as in (2). Now we consider the small-noise limit of (13). By the Freidlin-Wentzell theory and the contraction principle (see e.g. [9, Theorem 5.6.3]), the process  $(\xi_\varepsilon)_{\varepsilon>0}$  satisfies a large-deviation principle with a rate functional  $I$  given by

$$I(h; \mathbf{x}, \mathbf{y}) = \inf_{\xi} \left\{ \frac{1}{2} \int_0^h |\xi^{(n)}(t)|^2 dt : \xi \in C^n([0, h], \mathbb{R}^d) \text{ satisfying (2)} \right\}. \tag{14}$$

Up to a multiplicative constant, this rate functional is exactly the cost function  $C_{n,h}(\mathbf{x}, \mathbf{y})$  defined in (1).

In [8] the authors considered a more general system where the right-hand sides are



general functions of variables

$$\begin{aligned}
d\xi &= F_1(t, \xi, \xi_2) dt \\
d\xi_2 &= F_2(t, \xi, \xi_2, \xi_3) dt \\
&\vdots \\
d\xi_i &= F_i(t, \xi, \xi_2, \dots, \xi_{i+1}) dt \\
&\vdots \\
d\xi_n &= F_n(t, \xi, \dots, \xi_n) dt + \sigma(t, \xi, \dots, \xi_n) dW(t).
\end{aligned} \tag{15}$$

Equation (12) is a special (linear/Gaussian) case of (15) with  $F_i = \xi_{i+1}$ , for  $i = 1, \dots, n-1$ ,  $F_n = 0$  and  $\sigma = \sqrt{\varepsilon}$ . The reference [8, Theorem 1.1] provides two sided Gaussian bounds for the fundamental the solution of the forward Kolmogorov equation associated to (15), thus generalizes Aronson's estimate for uniformly elliptic diffusion processes [1] to a general hypo-elliptic setting. The rate functional  $I$  above (and hence the cost function  $C_{n,h}$  of the present paper) plays a key role in [8] because of two reasons:

- (1) It is related to the fundamental solution of the forward Kolmogorov equation associated to the Gaussian system which is proved using Fleming's logarithmic transform and control theory [8, Proposition 3.1].
- (2) [8, Theorem 1.1] is first proved for the Gaussian case using (1) then extended to the nonlinear case by linearizing.

In addition, in the case  $n = 2$  the cost function  $C_{2,h}$  (particularly its explicit formulation) has also been used to construct variational approximation schemes for the Kramer equation [22, 11], the system of isentropic Euler equations [18, 28] and the compressible Euler equations [6].

Based on the representation obtained in Theorem 1.2 of the present paper, in a companion paper [13] we provide an elementary proof for [8, Proposition 3.1] and extend [22, 11] to the forward Kolmogorov equation associated to (12) of which the Kramers equation is a special case.

### 3 Analytical formula of the cost functions

In this section, we prove Theorem 1.2 on the explicit formula for the cost function. Throughout the rest of the paper, all indices are numbered starting with zero.

*Proof. of Theorem 1.2.* We first recall the definition of a Wronskian matrix that will be used at several places later on. The Wronskian matrix  $W(f_1, \dots, f_n)$  associated to  $n$

functions  $f_1, \dots, f_n$  of a single real variable (say time) in the class  $C^n$  is defined by<sup>1</sup>

$$W(f_1, \dots, f_n) = \begin{pmatrix} f_1 & \dots & f_n \\ f_1' & \dots & f_n' \\ \vdots & \vdots & \vdots \\ f_1^{(n-1)} & \dots & f_n^{(n-1)} \end{pmatrix},$$

so that the  $(i, j)$ <sup>th</sup>-entry of this matrix is  $f_j^{(i)}$ , which is the  $i$ <sup>th</sup>-order derivative of  $f_j$ .

The optimal curve  $\xi$  in the definition of the cost function  $C_{n,h}$  satisfies the Euler-Lagrange equation

$$\xi^{(2n)}(t) = 0. \quad (16)$$

Therefore, it is a polynomial of order  $2n - 1$

$$\xi(t) = \sum_{i=0}^{2n-1} a_i t^i. \quad (17)$$

The coefficients  $\{a_i\}_{i=0}^{2n-1}$  will be determined from the boundary conditions. The  $k$ <sup>th</sup>-order derivative of  $\xi$  can be easily computed as

$$\begin{aligned} \xi^{(k)}(t) &= \sum_{i=k}^{2n-1} k! \binom{i}{k} a_i t^{i-k} \\ &= \begin{pmatrix} 0 & \dots & \underbrace{0}_{(k-1)\text{th entry}} & k! & \dots & \underbrace{k! \binom{i}{k} t^{i-k}}_{i\text{th entry, } k \leq i \leq n-1} & \dots & k! \binom{n-1}{k} t^{n-1-k} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} + \\ &+ \begin{pmatrix} k! \binom{n}{k} t^{n-k} & \dots & \underbrace{k! \binom{i}{k} t^{i-k}}_{(i-n+1)\text{th entry, } n \leq i \leq 2n-1} & \dots & k! \binom{2n-1}{k} t^{2n-1-k} \end{pmatrix} \begin{pmatrix} a_n \\ a_{n+1} \\ \vdots \\ a_{2n-1} \end{pmatrix}. \end{aligned}$$

Writing these equations for  $k = 0, \dots, n - 1$  in matrix form, we obtain

$$\begin{pmatrix} \xi(t) \\ \xi'(t) \\ \vdots \\ \xi^{(n-1)}(t) \end{pmatrix} = V_n \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} + A_n \begin{pmatrix} a_n \\ a_{n+1} \\ \vdots \\ a_{2n-1} \end{pmatrix}, \quad (18)$$

---

<sup>1</sup>Note that other authors sometimes denote by  $W(f_1, \dots, f_n)$  the determinant of the Wronskian matrix defined here.

where the matrices  $V_n$  and  $A_n$  depend on  $t$  and are given by

$$V_n(t) = \begin{bmatrix} 1 & \binom{2}{0}t & \binom{3}{0}t^2 & \cdots & \binom{k}{0}t^{k-1} & \cdots & \binom{n-1}{0}t^{n-2} \\ 0 & 1! & 1!\binom{3}{1}t^1 & \cdots & 1!\binom{k}{1}t^{k-1} & \cdots & 1!\binom{n-1}{1}t^{n-3} \\ 0 & 0 & 2! & \cdots & \vdots & \cdots & 2!\binom{n-1}{2}t^{n-4} \\ \vdots & \vdots & \cdots & \ddots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \cdots & k! & \cdots & k!\binom{n-1}{k}t^{n-1-k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 0 & (n-1)! \end{bmatrix}$$

and

$$A_n(t) = \begin{bmatrix} t^n & t^{n+1} & \cdots & t^{2n-1} \\ \binom{n}{1}t^{n-1} & \binom{n+1}{1}t^n & \cdots & \binom{2n-1}{1}t^{2n-2} \\ \vdots & \vdots & \vdots & \vdots \\ k!\binom{n}{k}t^{n-k} & k!\binom{n+1}{k}t^{n+1-k} & \cdots & k!\binom{2n-1}{k}t^{2n-1-k} \\ \vdots & \vdots & \vdots & \vdots \\ (n-1)!\binom{n}{n-1}t & (n-1)!\binom{n+1}{n-1}t^2 & \cdots & (n-1)!\binom{2n-1}{n-1}t^n \end{bmatrix}. \quad (19)$$

Note that  $V_n(t)$  and  $A_n(t)$  can be written in compact forms using the notation of the Wronskians

$$V_n(t) = W(1, t, \dots, t^{n-1}), \quad \text{and} \quad A_n(t) = W(t^n, \dots, t^{2n-1}).$$

In particular, when  $t = 0$ ,  $V_n(0) = \text{diag}(1, 1!, 2!, \dots, (n-1)!)$  is the diagonal matrix, and  $A_n(0) = 0$ . It follows that

$$\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} = V_n^{-1}(0) \begin{pmatrix} \xi(0) \\ \xi'(0) \\ \vdots \\ \xi^{(n-1)}(0) \end{pmatrix} = \begin{pmatrix} \xi(0) \\ \frac{1}{1!}\xi'(0) \\ \vdots \\ \frac{1}{(n-1)!}\xi^{(n-1)}(0) \end{pmatrix}.$$

Similarly, when  $t = h$ , we obtain

$$\begin{pmatrix} \xi(h) \\ \xi'(h) \\ \vdots \\ \xi^{(n-1)}(h) \end{pmatrix} = V_n(h) \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} + A_n(h) \begin{pmatrix} a_n \\ a_{n+1} \\ \vdots \\ a_{2n-1} \end{pmatrix}.$$

Therefore, we have the following equation to define  $a_n, \dots, a_{2n-1}$ ,

$$A_n(h) \begin{pmatrix} a_n \\ a_{n+1} \\ \vdots \\ a_{2n-1} \end{pmatrix} = \mathbf{b}_n,$$

where the vector  $\mathbf{b}_n$  on the right-hand side is given by

$$\mathbf{b}_n(h) = \begin{pmatrix} \xi(h) \\ \xi'(h) \\ \vdots \\ \xi^{(n-1)}(h) \end{pmatrix} - V_n(h) \begin{pmatrix} \xi(0) \\ \frac{1}{1!}\xi'(0) \\ \vdots \\ \frac{1}{(n-1)!}\xi^{(n-1)}(0) \end{pmatrix}.$$

The  $i$ -component of  $\mathbf{b}_n(h)$  can be computed explicitly using the definition of  $V_n(h)$  as follows

$$\begin{aligned} \mathbf{b}_n(h)[i] &= \xi^{(i)}(h) - \sum_{j=i}^{n-1} i! \binom{j}{i} h^{j-i} \frac{1}{j!} \xi^{(j)}(0) \\ &= \xi^{(i)}(h) - \sum_{j=i}^{n-1} i! \frac{j!}{(j-i)!i!} h^{j-i} \frac{1}{j!} \xi^{(j)}(0) \\ &= \xi^{(i)}(h) - \sum_{j=i}^{n-1} \frac{1}{(j-i)!} h^{j-i} \xi^{(j)}(0) \\ &= y_i - \sum_{j=i}^{n-1} \frac{1}{(j-i)!} h^{j-i} x_j. \end{aligned} \tag{20}$$

To proceed, we use the following lemma, whose proof is given below.

**Lemma 3.1.** *The matrix  $A_n(h)$  is invertible.*

Therefore, we can compute the coefficients  $a_n, \dots, a_{2n-1}$  from the matrix  $A_n(h)$  and the vector  $\mathbf{b}_n$ .

$$\begin{pmatrix} a_n \\ a_{n+1} \\ \vdots \\ a_{2n-1} \end{pmatrix} = A_n(h)^{-1} \mathbf{b}_n(h).$$

On the other hand, by integrating by parts successively, we obtain

$$\begin{aligned}
\int_0^h |\xi^{(n)}(t)|^2 dt &= \int_0^h \xi^{(n)}(t)^2 dt \\
&= \int_0^h \xi^{(n)}(t) d\xi^{(n-1)}(t) \\
&= \xi^{(n)}(t)\xi^{(n-1)}(t) \Big|_0^h - \int_0^h \xi^{(n-1)}(t) d\xi^{(n)}(t) \\
&= \xi^{(n)}(t)\xi^{(n-1)}(t) \Big|_0^h - \int_0^h \xi^{(n-1)}(t)\xi^{(n+1)}(t) dt \\
&= \xi^{(n)}(t)\xi^{(n-1)}(t) \Big|_0^h - \int_0^h \xi^{(n+1)}(t) d\xi^{(n-2)}(t) \\
&= \left( \xi^{(n)}(t)\xi^{(n-1)}(t) - \xi^{(n+1)}(t)\xi^{(n-2)}(t) \right) \Big|_0^h + \int_0^h \xi^{(n-2)}(t) d\xi^{(n+1)}(t) \\
&= \left( \xi^{(n)}(t)\xi^{(n-1)}(t) - \xi^{(n+1)}(t)\xi^{(n-2)}(t) \right) \Big|_0^h + \int_0^h \xi^{(n-2)}(t)\xi^{(n+2)}(t) dt \\
&= \dots \\
&= \sum_{i=0}^{n-1} (-1)^i \xi^{(n+i)}(t)\xi^{(n-1-i)}(t) \Big|_0^h + \int_0^h \xi(t)\xi^{(2n)}(t) dt \\
&= \sum_{i=0}^{n-1} (-1)^i \xi^{(n+i)}(t)\xi^{(n-1-i)}(t) \Big|_0^h, \tag{21}
\end{aligned}$$

where we have used the fact that  $\xi^{(2n)}(t) = 0$  to obtain the last equality. Next we will compute the last expression using the relation (18). It follows from (18) that

$$\begin{pmatrix} \xi^{(n)}(t) \\ \xi^{(n+1)}(t) \\ \vdots \\ \xi^{(n+k)}(t) \\ \vdots \\ \xi^{(2n-1)}(t) \end{pmatrix} = A_n^{(n)}(t) \begin{pmatrix} a_n \\ a_{n+1} \\ \vdots \\ a_k \\ \vdots \\ a_{2n-1} \end{pmatrix}$$

where  $A_n^{(n)}(t)$  is the matrix obtained from  $A_n(t)$  by taking  $n^{\text{th}}$ -order derivative of each entry of  $A_n(t)$ . The  $(k, i)^{\text{th}}$ -element,  $i = 1, \dots, n, k = 1, \dots, n, i \geq k$ , of  $A_n^{(n)}(t)$  is given by

$$\begin{aligned}
\left( (k-1)! \binom{n+i-1}{k-1} t^{n+i-k} \right)^{(n)} &= (k-1)! \binom{n+i-1}{k-1} n! \binom{i-k+n}{i-k} t^{-k+i} \\
&= \frac{(n+i-1)!}{(i-k)!} t^{i-k}.
\end{aligned}$$

Other elements of  $A_n^{(n)}(t)$  are equal to 0. Therefore, we have

$$\begin{pmatrix} \xi^{(n)}(t)\xi^{(n-1)}(t) \\ \xi^{(n+1)}(t)\xi^{(n-2)}(t) \\ \vdots \\ \xi^{(n+k)}(t)\xi^{(n-k-1)}(t) \\ \vdots \\ \xi^{(2n-1)}(t)\xi(t) \end{pmatrix} = D_n(t) \begin{pmatrix} a_n \\ a_{n+1} \\ \vdots \\ a_k \\ \vdots \\ a_{2n-1} \end{pmatrix},$$

where the  $(k, i)^{\text{th}}$ -element,  $i = 1, \dots, n, k = 1, \dots, n, i \geq k$ , of  $D_n(t)$  is given by  $\frac{(n+i-1)!}{(i-k)!} t^{i-k} \xi^{(n-k)}(t)$ . Other elements of  $D_n(t)$  are equal to 0. It follows that

$$\begin{pmatrix} \xi^{(n)}(h)\xi^{(n-1)}(h) \\ \xi^{(n+1)}(h)\xi^{(n-2)}(h) \\ \vdots \\ \xi^{(n+k)}(h)\xi^{(n-k-1)}(h) \\ \vdots \\ \xi^{(2n-1)}(h)\xi(h) \end{pmatrix} - \begin{pmatrix} \xi^{(n)}(0)\xi^{(n-1)}(0) \\ \xi^{(n+1)}(0)\xi^{(n-2)}(0) \\ \vdots \\ \xi^{(n+k)}(0)\xi^{(n-k-1)}(0) \\ \vdots \\ \xi^{(2n-1)}(0)\xi(0) \end{pmatrix} = E_n(h)[A_n(h)]^{-1}\mathbf{b}_n(h),$$

where the matrix  $E_n(h) = D_n(h) - D_n(0)$  is given by

$$E_n(h)[i_1, i_2] = \begin{cases} \frac{(n+i_2)!}{(i_2-i_1)!} \xi^{(n-i_1-1)}(h) h^{-i_1+i_2}, & i_2 > i_1, \\ \frac{(n+i_2)!}{(i_2-i_1)!} [\xi^{(n-i_1-1)}(h) - \xi^{(n-i_1-1)}(0)], & i_2 = i_1, \\ 0, & i_2 < i_1, \end{cases}$$

for all  $i_1, i_2 = 0, \dots, n$ . Therefore, substituting this back to (21), we obtain

$$C_h = \int_0^h (\xi^{(n)}(t))^2 dt = \begin{pmatrix} 1 & -1 & 1 & -1 & \dots \end{pmatrix} E_n(h)[A_n(h)]^{-1}\mathbf{b}_n(h). \quad (22)$$

The right-hand side of (22) can be transformed further using the following lemma, whose proof is presented below.

**Lemma 3.2.** *It holds that*

$$\begin{bmatrix} 1 & -1 & 1 & -1 & \dots \end{bmatrix} E_n(h) = [\mathbf{b}_n(h)]^T B_n(h), \quad (23)$$

where the matrix  $B_n(h)$  is defined as follows

$$B_n(h)[i_1, i_2] = \begin{cases} (-1)^{n-i_1-1} \frac{(n+i_2)!}{(i_1+i_2-n+1)!} h^{i_2+i_1-n+1}, & i_2 + i_1 \geq n-1, \\ 0 & i_2 + i_1 < n-1, \end{cases} \quad (24)$$

for all  $i_1, i_2 = 0, \dots, n-1$ .

Substituting (23) into (22), we obtain

$$C_h = \int_0^h (\xi^{(n)}(t))^2 dt = [\mathbf{b}_n(h)]^T B_n(h) [A_n(h)]^{-1} \mathbf{b}_n(h),$$

with  $\mathbf{b}_n(h)$ ,  $A_n(h)$  and  $B_n(h)$  defined in (20), (19) and (24) respectively. This establishes the statement of Theorem 1.2.  $\square$

For completion, we now prove Lemma 3.1 and Lemma 3.2.

*Proof. of Lemma 3.1.* We recall that  $A_n(h)$  can be written in terms of the Wronskian

$$A_n(h) = W(h^n, \dots, h^{2n-1}).$$

Therefore

$$\det A_n(h) = \det W(f_1, \dots, f_n).$$

According to [3, Lemma 1], we have

$$\det A_n(h) = V(n, \dots, 2n-1) h^{\sum_{i=n}^{2n-1} i - \binom{n}{2}} = V(n, \dots, 2n-1) h^{\frac{n(3n-1)}{2}},$$

where  $V(n, \dots, 2n-1)$  is the Vandermonde determinant

$$V(n, \dots, 2n-1) = \begin{vmatrix} 1 & \dots & 1 \\ n & \dots & 2n-1 \\ \vdots & \vdots & \vdots \\ n^{n-1} & \dots & (2n-1)^{n-1} \end{vmatrix} = \prod_{1 \leq i < j \leq n} (j-i).$$

Hence, we obtain

$$\det A_n(h) = h^{\frac{n(3n-1)}{2}} \prod_{1 \leq i < j \leq n} (j-i), \quad (25)$$

which is non-zero. This completes the proof of the lemma.  $\square$

*Proof. of Lemma 3.2.* The equality (23) is interesting on its own. Below we will prove it using purely combinatorial techniques.

The  $k^{\text{th}}$  element on the left hand side of (23) is

$$\sum_{i=0}^{n-1} (-1)^i E_n(h)[i, k] = (n+k)! (-1)^k (\xi^{(n-k-1)}(h) - \xi^{(n-k-1)}(0)) + \sum_{i=0}^{k-1} (-1)^i \frac{(n+k)!}{(k-i)!} \xi^{(n-i-1)}(h) h^{-i+k}.$$

The  $(i_1, i_2)$  element of  $B_n(h)$  is

$$B_n(h)[i_1, i_2] = \begin{cases} (-1)^{n-i_1-1} \frac{(n+i_2)!}{(i_1+i_2-n+1)!} h^{i_2+i_1-n+1}, & i_2 + i_1 \geq n-1, \\ 0 & i_2 + i_1 < n-1. \end{cases}$$

So that the  $k^{\text{th}}$  element on the right hand side of (23) is

$$\begin{aligned}
\sum_{i=0}^{n-1} \mathbf{b}_n(i) B_n[i, k] &= \sum_{i=0, i+k \geq n-1}^{n-1} \mathbf{b}_n(i) (-1)^{n-i-1} \frac{(n+k)!}{(k+i-n+1)!} h^{k+i-n+1} \\
&= \sum_{i=n-1-k}^{n-1} \mathbf{b}_n(i) (-1)^{n-i-1} \frac{(n+k)!}{(k+i-n+1)!} h^{k+i-n+1} \\
&= \sum_{i=n-1-k}^{n-1} [\xi^{(i)}(h) - \sum_{j=i}^{n-1} \frac{1}{(j-i)!} h^{j-i} \xi^{(j)}(0)] (-1)^{n-i-1} \frac{(n+k)!}{(k+i-n+1)!} h^{k+i-n+1}.
\end{aligned}$$

To establish (23), we need to show that

$$\begin{aligned}
&\sum_{i=n-1-k}^{n-1} [\xi^{(i)}(h) - \sum_{j=i}^{n-1} \frac{1}{(j-i)!} h^{j-i} \xi^{(j)}(0)] (-1)^{n-i-1} \frac{(n+k)!}{(k+i-n+1)!} h^{k+i-n+1} \\
&= (n+k)! (-1)^k (\xi^{(n-k-1)}(h) - \xi^{(n-k-1)}(0)) + \sum_{i=0}^{k-1} (-1)^i \frac{(n+k)!}{(k-i)!} \xi^{(n-i-1)}(h) h^{-i+k}. \quad (26)
\end{aligned}$$

We will show this by transforming the left-hand side. First we write it as follows

$$\begin{aligned}
&\sum_{i=n-1-k}^{n-1} [\xi^{(i)}(h) - \sum_{j=i}^{n-1} \frac{1}{(j-i)!} h^{j-i} \xi^{(j)}(0)] (-1)^{n-i-1} \frac{(n+k)!}{(k+i-n+1)!} h^{k+i-n+1} \\
&= \sum_{i=n-1-k}^{n-1} \xi^{(i)}(h) (-1)^{n-i-1} \frac{1}{(k+i-n+1)!} h^{k+i-n+1} \\
&\quad - \sum_{i=n-1-k}^{n-1} \sum_{j=i}^{n-1} \frac{1}{(j-i)!} h^{j-i} \xi^{(j)}(0) (-1)^{n-i-1} \frac{1}{(k+i-n+1)!} h^{k+i-n+1} \\
&= (I) - (II).
\end{aligned}$$

Now we transform further (I) and (II). We have

$$\begin{aligned}
(I) &:= \sum_{i=n-1-k}^{n-1} \xi^{(i)}(h) (-1)^{n-i-1} \frac{1}{(k+i-n+1)!} h^{k+i-n+1} \\
&= \sum_{i=n-1-k}^{n-1} \xi^{(i)}(h) (-1)^{n-i-1} \frac{1}{(k+i-n+1)!} h^{k+i-n+1} \quad (\text{now we change variable: } i = -j + n - 1) \\
&= \sum_{i=k}^0 \xi^{(n-j-1)}(h) (-1)^j \frac{1}{(k-j)!} h^{k-j} \\
&= \xi^{(n-k-1)}(h) (-1)^k + \sum_{i=0}^{k-1} \xi^{(n-j-1)}(h) (-1)^j \frac{1}{(k-j)!} h^{k-j}. \quad (27)
\end{aligned}$$



The term (II) can be transformed as follows.

$$\begin{aligned}
(II) &:= - \sum_{i=n-1-k}^{n-1} \sum_{j=i}^{n-1} \frac{1}{(j-i)!} h^{j-i} \xi^{(j)}(0) (-1)^{n-i-1} \frac{1}{(k+i-n+1)!} h^{k+i-n+1} \\
&= \sum_{i=n-1-k}^{n-1} \sum_{j=i}^{n-1} \frac{1}{(j-i)!} h^{j-i} \xi^{(j)}(0) (-1)^{n-i} \frac{1}{(k+i-n+1)!} h^{k+i-n+1} \\
&= \sum_{j=n-1-k}^{n-1} \xi^{(j)}(0) h^{j+k-n+1} \sum_{i=n-1-k}^j \frac{1}{(j-i)!} (-1)^{n-i} \frac{1}{(k+i-n+1)!} \quad (\text{for } j = n-1-k..n-1) \\
&= -\xi^{(n-k-1)}(0) (-1)^k + \sum_{j=n-k}^{n-1} \xi^{(j)}(0) h^{j+k-n+1} \sum_{i=n-1-k}^j \frac{1}{(j-i)!} (-1)^{n-i} \frac{1}{(k+i-n+1)!}.
\end{aligned} \tag{28}$$

We now show that the second term of (28) vanishes by showing that

$$\sum_{i=n-1-k}^j \frac{1}{(j-i)!} (-1)^{n-i} \frac{1}{(k+i-n+1)!} = 0$$

for each  $j = n-k, \dots, n-1$ . In fact, we have

$$\begin{aligned}
&\sum_{i=n-1-k}^j (-1)^{n-i} \frac{1}{(j-i)! (k+i-n+1)!} \\
&= \frac{1}{(j+k-n+1)!} \sum_{i=n-1-k}^j (-1)^{n-i} \frac{(j+k-n+1)!}{(j-i)! (k+i-n+1)!} \\
&= \frac{(-1)^{n-j}}{(j+k-n+1)!} \sum_{i=n-1-k}^j (-1)^{j-i} \binom{j+k-n+1}{j-i} \\
&= \frac{(-1)^{n-j}}{(j+k-n+1)!} \sum_{l=j-(n-1-k)}^0 (-1)^l \binom{j+k-n+1}{l} \quad (\text{now we change variable: } l = j-i, i = j-l) \\
&= \frac{(-1)^{n-j}}{(j+k-n+1)!} \sum_{l=0}^{j-n+1+k} (-1)^l \binom{j+k-n+1}{l} \\
&= \frac{(-1)^{n-j}}{(j+k-n+1)!} (-1+1)^{j-n+k+1} \\
&= 0.
\end{aligned}$$

As a consequence,  $(II) = -\xi^{(n-k-1)}(0) (-1)^k$ . Therefore, the left-hand side of (26) is equal to

$$(\xi^{(n-k-1)}(h) - \xi^{(n-k-1)}(0)) (-1)^k + \sum_{i=0}^{k-1} \xi^{(n-j-1)}(h) (-1)^j \frac{1}{(k-j)!} h^{k-j},$$

which is exactly its right-hand side. This finishes the proof of Lemma 3.2.  $\square$

**Remark 3.3.** We recall that the cost function  $C_{n,h}$  is given by

$$C_{n,h} = \mathbf{b}_n(h)^T B_n(h) A_n^{-1}(h) \mathbf{b}_n(h). \quad (29)$$

We can show that the matrix  $B_n(h) A_n^{-1}(h)$  is positive definite. Indeed, let  $z \in \mathbb{R}^{dn}$  be arbitrary. Then by definition of  $C_{n,h}$  and from (29), we have

$$z^T B_n(h) A_n^{-1}(h) z = C_{n,h}(0, z).$$

Then the positive definiteness of  $B_n(h) A_n^{-1}(h)$  follows from Lemma 2.2. Note further that we have the following property, for any matrix  $M$  then

$$z^T M z = (z^T M z)^T = z^T M^T z = z^T \frac{M + M^T}{2} z. \quad (30)$$

Therefore we can write

$$C_{n,h} = \mathbf{b}_n(h)^T H_n(h) \mathbf{b}_n(h),$$

where  $H_n(h) = \frac{1}{2} [B_n(h) A_n^{-1}(h) + (B_n(h) A_n^{-1}(h))^T]$  is a symmetric positive matrix.

**Remark 3.4** (Alternative computations). We can compute the cost function  $C_{n,h}$  more directly as follows.

$$\begin{aligned} \int_0^h |\xi^{(n)}(t)|^2 dt &= \int_0^h \left( \sum_{i=n}^{2n-1} n! \binom{i}{n} a_i t^{i-n} \right)^2 dt \\ &= (n!)^2 \sum_{n \leq i, j \leq 2n-1} \binom{i}{n} \binom{j}{n} a_i a_j \int_0^h t^{i+j-2n} dt \\ &= (n!)^2 \sum_{n \leq i, j \leq 2n-1} \binom{i}{n} \binom{j}{n} a_i a_j \frac{h^{i+j+1-2n}}{i+j+1-2n}. \end{aligned}$$

Define the matrix  $K_n(h) = (K_{ij})_{i,j=n}^{2n-1}$  with entries

$$K_{ij} = (n!)^2 \binom{i}{n} \binom{j}{n} \frac{h^{i+j+1-2n}}{i+j+1-2n}.$$

Clearly  $K_n(h)$  is symmetric. Then  $C_{n,h}$  can be written as

$$C_{n,h} = a^T K_n(h) a = \mathbf{b}_n^T(A_n(h)^{-1})^T K_n(h) A_n(h)^{-1} \mathbf{b}_n(h),$$

where  $a = (a_n, \dots, a_{2n-1})^T$  and  $\mathbf{b}_n(h)$  and  $A_n(h)$  are defined in (24) and (19) respectively.

The advantage of the formula derived in the previous section is that it involves  $A_n^{-1}$  only one time and the matrix  $B_n$  is triangular.

**Remark 3.5** (Alternative representation using the scaling property). Using the scaling property in Theorem 2.1, the cost function can be written as follows

$$C_{n,h} = h^{1-2n} [\tilde{\mathbf{b}}_n(h)]^T B_n(1) [A_n(1)]^{-1} \tilde{\mathbf{b}}_n(h),$$

where

$$\tilde{\mathbf{b}}_n(h) = \begin{pmatrix} \tilde{\xi}(1) \\ h\tilde{\xi}'(1) \\ \vdots \\ h^k\tilde{\xi}^{(k)}(1) \\ \vdots \\ h^{n-1}\tilde{\xi}^{(n-1)}(1) \end{pmatrix} - V_n(1) \begin{pmatrix} \tilde{\xi}(0) \\ h\tilde{\xi}'(0) \\ \vdots \\ h^k\tilde{\xi}^{(k)}(0) \\ \vdots \\ h^{n-1}\tilde{\xi}^{(n-1)}(0) \end{pmatrix}.$$

By analogous computations to (20), the  $i$ -component of this vector is

$$\tilde{\mathbf{b}}_n(h)[i] = y_i h^i - \sum_{j=i}^{n-1} \frac{1}{(j-i)!} h^j x_j = h^i \mathbf{b}_n(h)[i].$$

## 4 LU decomposition of $A_n$ and $A_n^{-1}$

The analytical formulas for the cost functions obtained in the previous section involve the inverse of the matrix  $A_n$ , which is a matrix of order  $n$ . In this section, we provide an explicit formula for the  $LU$  decomposition of  $A_n$  and for  $A_n^{-1}$ . For simplicity of notation, we leave out the dependence on  $h$  of  $A_n$  (and hence  $L, U$ ). We recall that  $A = W(h^n, \dots, h^{2n-1})$  is the Wronskian matrix associated to the polynomials  $\{h^n, \dots, h^{2n-1}\}$ , therefore the analysis of this section is of independent interest since the Wronskian matrix plays an important role in linear algebra and differential equations.

The main result of this section is the following theorem, which is summarised as Theorem 1.3 in the introduction.

**Theorem 4.1.** 1.  $A_n = LU$  where  $U$  and  $L$  are defined as follows

$$U[i, j] = \begin{cases} \frac{(j-1)!}{(j-i)!} h^{j+n-i} & \text{if } j \geq i, \\ 0 & \text{otherwise,} \end{cases} \quad (31)$$

and

$$L[k, j] = \begin{cases} h^{j-k} \binom{k-1}{j-1} \frac{n!}{(n-k+j)!} & \text{if } j \leq k, \\ 0 & \text{otherwise.} \end{cases} \quad (32)$$

2. The inverse of  $A_n$  is given by the product of the following two matrices:

$$U^{-1}[i, j] = \begin{cases} \frac{(-1)^{i+j}}{((i-1)!(j-i)!h^{-j+i+n})} & \text{if } j \geq i, \\ 0 & \text{otherwise,} \end{cases} \quad (33)$$

and

$$\begin{aligned}
L^{-1}[j, i] &= \begin{cases} (-1)^{j-i} h^{i-j} \frac{(j-1)!}{(j-i)!(i-1)!} \frac{(n+j-i-1)!}{(n-1)!} & \text{if } j \geq i, \\ 0 & \text{otherwise,} \end{cases} \\
&= \begin{cases} (-1)^{j-i} h^{i-j} \frac{(j-1)!}{(i-1)!} \binom{n+j-i-1}{j-i} & \text{if } j \geq i \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{34}$$

*Proof.* We first prove the first statement about the  $LU$  decomposition of the matrix  $A_n$ . We will show that  $U^{-1}$  can be computed as follows:

$$U^{-1}[i, j] = \begin{cases} \frac{1}{((i-1)!(j-i)!)h^{-j+i+n}} & \text{if } j \geq i, \\ 0 & \text{otherwise.} \end{cases}$$

In fact, we have

$$\begin{aligned}
\sum_j U[i, j] U^{-1}(j, k) &= \sum_{j \geq i, k \geq j} U[i, j] U^{-1}[j, k] \\
&= \sum_{j=i}^k U[i, j] U^{-1}[j, k] \\
&= \sum_{j=i}^k \frac{(j-1)!}{(j-i)!} h^{j+n-i} \frac{(-1)^{j+k}}{((j-1)!(k-j)!)h^{n-k+j}} \\
&= h^{k-i} (-1)^k \sum_{j=i}^k \frac{1}{(j-i)!} \frac{(-1)^j}{(k-j)!}.
\end{aligned}$$

If  $k = i$ , then  $\sum_j U[i, j] U^{-1}[j, k] = (-1)^k \frac{1}{(k-i)!} \frac{(-1)^k}{(k-k)!} = 1$ .

If  $k \neq i$ , then we have

$$\begin{aligned}
\sum_{j=i}^k \frac{1}{(j-i)!} \frac{(-1)^j}{(k-j)!} &= \frac{1}{(k-i)!} \sum_{j=i}^k \frac{(k-i)!}{(j-i)!(k-j)!} (-1)^j \\
&= \frac{1}{(k-i)!} (-1)^i \sum_{l=0}^{k-i} \frac{(k-i)!}{l!(k-i-l)!} (-1)^{l+i} \quad (\text{change variable: } l = j - i) \\
&= \frac{1}{(k-i)!} (-1)^i (1 + (-1))^{k-i} \\
&= 0.
\end{aligned}$$

Therefore,

$$\sum_j U[i, j] U^{-1}[j, k] = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{if } k \neq i. \end{cases}$$

In other words,  $U^{-1}$  can be defined as in Eq. (33).

Now we establish the formula for  $L$ . By definition  $A = LU$ , so that  $L[k, j]$  can be written as follows

$$L[k, j] = \sum_{i \leq j} A_n[k, i] U^{-1}[i, j] = h^{j-k} \sum_{i=1}^j \frac{(n+i-1)!}{(n+i-k)!} \frac{(-1)^{i+j}}{((i-1)!(j-i)!)}.$$

We now simplify the expression above. Consider the function  $f(x) = x^n(1-x)^{j-1}$ . On the one hand, we have

$$\begin{aligned} f(x) &= x^n \sum_{i=0}^{j-1} \binom{j-1}{i} (-1)^i x^i \\ &= \sum_{i=0}^{j-1} \binom{j-1}{i} (-1)^i x^{n+i} \\ &= \sum_{i=1}^j \binom{j-1}{i-1} (-1)^{i-1} x^{n+i-1}. \end{aligned}$$

Therefore,

$$\begin{aligned} f(x)^{(k-1)} &= \sum_{i=1}^j \binom{j-1}{i-1} (-1)^{i-1} \frac{(n+i-1)!}{(n+i-1-(k-1))!} x^{n+i-1-(k-1)} \\ &= \sum_{i=1}^j \frac{(j-1)!}{(j-i)!(i-1)!} (-1)^{i-1} \frac{(n+i-1)!}{(n+i-k)!} x^{n+i-1-(k-1)}. \end{aligned}$$

It follows that

$$\sum_{i=1}^j \frac{(-1)^i}{(i-1)!(j-i)!} \frac{(n+i-1)!}{(n+i-k)!} = \frac{-1}{(j-1)!} f(1)^{(k-1)},$$

and by multiplying both sides of this equality with  $h^{j-k}$ , we get

$$h^{j-k} \sum_{i=1}^j \frac{(-1)^{i+j}}{(i-1)!(j-i)!} \frac{(n+i-1)!}{(n+i-k)!} = h^{j-k} \frac{(-1)^{j+1}}{(j-1)!} f(1)^{(k-1)}.$$

On the other hand, according to Leibniz formula, we have

$$\begin{aligned} f(x)^{(k-1)} &= [x^n(1-x)^{j-1}]^{(k-1)} \\ &= \sum_{i=0}^{k-1} \binom{k-1}{i} [x^n]^{(k-1-i)} [(1-x)^{j-1}]^{(i)}. \end{aligned}$$

If  $k-1 < j-1$ , then  $f(1)^{(k-1)} = 0$ , which implies that  $\sum_{i \leq j} A_n[k, i] U^{-1}[i, j] = 0$ .

If  $k \geq j$ , then we have

$$\begin{aligned}
f(x)^{(k-1)}|_{x=1} &= \sum_{i=0}^{k-1} \binom{k-1}{i} [x^n]^{(k-1-i)} [(1-x)^{j-1}]^{(i)}|_{x=1} \\
&= \sum_{i=0}^{j-1} \binom{k-1}{i} [x^n]^{(k-1-i)} [(1-x)^{j-1}]^{(i)}|_{x=1} \\
&\quad + \sum_{i=j}^{k-1} \binom{k-1}{i} [x^n]^{(k-1-i)} [(1-x)^{j-1}]^{(i)}|_{x=1} \\
&= \binom{k-1}{j-1} [x^n]^{(k-1-(j-1))} (-1)^{j-1} [(1-x)^{j-1}]^{(j-1)}|_{x=1} \\
&\quad + 0 \\
&= \binom{k-1}{j-1} \frac{n!}{(n-(k-j))!} x^{n-(k-j)} (j-1)! (-1)^{j-1}|_{x=1} \\
&= \binom{k-1}{j-1} \frac{n!}{(n-(k-j))!} (j-1)! (-1)^{j-1}.
\end{aligned}$$

Therefore,

$$h^{j-k} \sum_{i=1}^j \frac{(-1)^{i+j}}{(i-1)!(j-i)!} \frac{(n+i-1)!}{(n+i-k)!} = h^{j-k} (-1)^{j+1} (-1)^{j-1} \binom{k-1}{j-1} \frac{n!}{(n-(k-j))!}.$$

It then follows that

$$L(k, j) = \sum_{i \leq j} A_n(k, i) U^{-1}(i, j) = h^{j-k} \binom{k-1}{j-1} \frac{n!}{(n-k+j)!},$$

which is the desired formula. This completes the proof of the first statement of the theorem.

Now we will prove the second statement of the theorem. We will show that  $L^{-1}$  has the form as defined in (34).

The element  $(k, i)$  of the product of  $L$  and  $L^{-1}$  is given by  $\sum_{j=1}^n L[k, j] L^{-1}[j, i]$ .

If  $k < i$ , then

$$\sum_{j=1}^n L[k, j] L^{-1}[j, i] = \sum_{j=1}^k L[k, j] L^{-1}[j, i] = 0. \quad (35)$$

If  $k > i$ , then we have

$$\begin{aligned}
\sum_{j=1}^n L[k, j] L^{-1}[j, i] &= \sum_{j=1, j \geq i, j \leq k}^n L[k, j] L^{-1}[j, i] \\
&= \sum_{j=i}^k h^{j-k} \binom{k-1}{j-1} \frac{n!}{(n-k+j)!} (-1)^{j-i} h^{i-j} \frac{(j-1)!}{(j-i)!(i-1)!} \frac{(n+j-i-1)!}{(n-1)!} \\
&= h^{i-k} \sum_{j=i}^k \frac{(k-1)!}{(k-j)!} \frac{n!}{(n-k+j)!} (-1)^{j-i} \frac{1}{(j-i)!(i-1)!} \frac{(n+j-i-1)!}{(n-1)!} \\
&= h^{i-k} \frac{(k-1)!n!}{(k-i)!(i-1)!(n-1)!} \sum_{j=0}^{k-i} \frac{(k-i)!}{(k-i-j)!j!} \frac{(n+j-1)!}{(n-k+i+j)!} (-1)^j,
\end{aligned}$$

where to obtain the last equality we have changed variable  $j := j + i$ . Next, we will show that the summation in the above expression is equal to 0.

Let  $\alpha = k - i$  and consider  $g(x) = x^{n-1}(x-1)^\alpha$ . On the one hand, we have

$$g(x) = \sum_{j=0}^{\alpha} \binom{\alpha}{j} x^{n+j-1} (-1)^{\alpha-j},$$

which follows that

$$\begin{aligned}
g^{(\alpha-1)}(x) &= \sum_{j=0}^{\alpha} \binom{\alpha}{j} [x^{n+j-1}]^{(\alpha-1)} (-1)^{\alpha-j} \\
&= \sum_{j=0}^{\alpha} \binom{\alpha}{j} \frac{(n+j-1)!}{(n+j-1-(\alpha-1))!} x^{n+j-\alpha} (-1)^{\alpha-j} \\
&= \sum_{j=0}^{\alpha} \binom{\alpha}{j} \frac{(n+j-1)!}{(n+j-\alpha)!} x^{n+j-\alpha} (-1)^{\alpha-j} \\
&= (-1)^\alpha \sum_{j=0}^{k-i} \frac{(k-i)!}{(k-i-j)!j!} \frac{(n+j-1)!}{(n+j-k+i)!} x^{n+j-\alpha} (-1)^j.
\end{aligned}$$

In particular, we obtain

$$g^{(\alpha-1)}(1) = (-1)^\alpha \sum_{j=0}^{k-i} \frac{(k-i)!}{(k-i-j)!j!} \frac{(n+j-1)!}{(n+j-k+i)!} (-1)^j.$$

On the other hand, we have

$$\begin{aligned}
g^{(\alpha-1)}(x) &= [x^{n-1}(x-1)^\alpha]^{(\alpha-1)} \\
&= \sum_{j=0}^{\alpha-1} [x^{n-1}]^{(\alpha-1-j)} [(x-1)^\alpha]^{(j)},
\end{aligned}$$

so  $g^{(\alpha-1)}(1) = 0$ .

Therefore,

$$(-1)^\alpha \sum_{j=0}^{k-i} \frac{(k-i)!}{(k-i-j)!j!} \frac{(n+j-1)!}{(n+j-k+i)!} (-1)^j = 0,$$

and hence we obtain for  $k > i$ ,

$$\sum_{j=1}^n L[k, j] L^{-1}[j, i] = 0. \quad (36)$$

Finally when  $k = i$ , we have

$$\begin{aligned} \sum_{j=1}^n L[k, j] L^{-1}[j, i] &= \sum_{j=1}^i L[k, j] L^{-1}[j, k] + \sum_{j=i+1}^n L[k, j] L^{-1}[j, k] \\ &= \sum_{j=1}^k L[k, j] L^{-1}[j, k] \\ &= L[k, k] L^{-1}[k, k] \\ &= 1, \end{aligned} \quad (37)$$

where in the last step we have used that  $L[k, k] = L^{-1}[k, k] = 1$ . From (35), (36), and (37), we conclude that (34) is indeed the formula for the inverse of  $L$ .

This finishes the proof of the theorem.  $\square$

## 5 Numerical investigations

In this section, we provide an algorithm to compute  $C_{n,h}$  using results from the previous section and compute the expressions obtained for small  $n$  explicitly.

**Algorithm 5.1.** *Algorithm to compute  $C_{n,h}(\mathbf{x}, \mathbf{y})$ :*

*Input:*  $n, h, \mathbf{x} = (x_0, \dots, x_{n-1}) \in \mathbb{R}^{dn}$  and  $\mathbf{y} = (y_0, \dots, y_{n-1}) \in \mathbb{R}^{dn}$ .

*Output:*  $C_{n,h}(\mathbf{x}, \mathbf{y})$ .

*The algorithm consists of 5 steps.*

*Step 1: Compute  $B_n(h)$*

$$B_n(h)[i_1, i_2] = \begin{cases} (-1)^{n-i_1-1} \frac{(n+i_2)!}{(i_1+i_2-n+1)!} h^{i_2+i_1-n+1}, & i_2 + i_1 \geq n-1, \\ 0 & i_2 + i_1 < n-1. \end{cases}$$

*Step 2: Compute  $\mathbf{b}_n(h)$*

$$\mathbf{b}_n(h)[i] = y_i - \sum_{j=i}^{n-1} \frac{1}{(j-i)!} h^{j-i} x_j.$$

*Step 3: Compute  $L_n^{-1}(h)$  and  $U_n^{-1}(h)$*



$$U^{-1}[i, j] = \begin{cases} \frac{(-1)^{i+j}}{((i-1)!(j-i)!h^{-j+i+n})} & \text{if } j \geq i, \\ 0 & \text{otherwise,} \end{cases}, \quad L^{-1}[j, i] = \begin{cases} (-1)^{j-i} h^{i-j} \frac{(j-1)!}{(i-1)!} \binom{n+j-i-1}{j-i} & \text{if } j \geq i \\ 0 & \text{otherwise} \end{cases}$$

*Step 4: Compute  $A_n^{-1} = U_n^{-1}L_n^{-1}$ .*

*Step 5: Compute  $C$  using Theorem 1.2*

$$C_{n,h}(x_0, \dots, x_{n-1}; y_0, \dots, y_{n-1}) = \mathbf{b}_n(h)^T B_n(h) [A_n(h)]^{-1} \mathbf{b}_n(h).$$

Below we show  $B_n(h), A_n(h), L_n(h), U_n(h), L_n^{-1}(h), U_n^{-1}(h)$  and  $C_{n,h}$  for  $n = 1, 2, 3, 4$  computed using Algorithm 5.1. For simplicity of notation, we leave out the dependence on  $h$  of the matrices.

**For  $n = 1$ .**

$$\begin{aligned} B &= \begin{pmatrix} 1 \end{pmatrix}, \\ A &= \begin{pmatrix} h \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} \frac{1}{h} \end{pmatrix}, \\ L &= \begin{pmatrix} 1 \end{pmatrix}, \quad U = \begin{pmatrix} h \end{pmatrix}, \\ L^{-1} &= \begin{pmatrix} 1 \end{pmatrix}, \quad U^{-1} = \begin{pmatrix} \frac{1}{h} \end{pmatrix}, \\ C_{1,h}(x_0, y_0) &= \frac{1}{h} (-x_0 + y_0)^2. \end{aligned}$$

**For  $n = 2$**

$$\begin{aligned} B &= \begin{pmatrix} 0 & -6 \\ 2 & 6h \end{pmatrix}, \\ A &= \begin{pmatrix} h^2 & h^3 \\ 2h & 3h^2 \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} \frac{3}{h^2} & -\frac{1}{h} \\ -\frac{2}{h^3} & \frac{1}{h^2} \end{pmatrix}, \\ L &= \begin{pmatrix} 1 & 0 \\ \frac{2}{h} & 1 \end{pmatrix}, \quad U = \begin{pmatrix} h^2 & h^3 \\ 0 & h^2 \end{pmatrix}, \\ L^{-1} &= \begin{pmatrix} 1 & 0 \\ -\frac{2}{h} & 1 \end{pmatrix}, \quad U^{-1} = \begin{pmatrix} \frac{1}{h^2} & -\frac{1}{h} \\ 0 & \frac{1}{h^2} \end{pmatrix}, \\ C_{2,h}(x_0, x_1; y_0, y_1) &= \frac{1}{h} \left[ (y_1 - x_1)^2 + 3 \left( y_1 - x_1 - \frac{2(y_0 - x_0 - hx_1)}{h} \right)^2 \right]. \end{aligned}$$

For  $n = 3$

$$\begin{aligned}
B &= \begin{pmatrix} 0 & 0 & 120 \\ 0 & -24 & -120h \\ 6 & 24h & 60h^2 \end{pmatrix}, \\
A &= \begin{pmatrix} h^3 & h^4 & h^5 \\ 3h^2 & 4h^3 & 5h^4 \\ 6h & 12h^2 & 20h^3 \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} \frac{10}{h^3} & -\frac{4}{h^2} & \frac{1}{2h} \\ -\frac{15}{h^4} & \frac{h^3}{h^3} & -\frac{1}{h^2} \\ \frac{6}{h^5} & -\frac{3}{h^4} & \frac{1}{2h^3} \end{pmatrix} \\
L &= \begin{pmatrix} 1 & 0 & 0 \\ \frac{3}{h} & 1 & 0 \\ \frac{6}{h^2} & \frac{6}{h} & 1 \end{pmatrix}, \quad U = \begin{pmatrix} h^3 & h^4 & h^5 \\ 0 & h^3 & 2h^4 \\ 0 & 0 & 2h^3 \end{pmatrix}, \\
L^{-1} &= \begin{pmatrix} 1 & 0 & 0 \\ -\frac{3}{h} & 1 & 0 \\ \frac{12}{h^2} & -\frac{6}{h} & 1 \end{pmatrix}, \quad U^{-1} = \begin{pmatrix} \frac{1}{h^3} & -\frac{1}{h^2} & \frac{1}{2h} \\ 0 & \frac{1}{h^3} & -\frac{1}{h^2} \\ 0 & 0 & \frac{1}{2h^3} \end{pmatrix},
\end{aligned}$$

and

$$\begin{aligned}
C_{3,h}(x_0, x_1, x_2; y_0, y_1, y_2) &= \frac{1}{h} \left[ (y_2 - x_2)^2 + 3 \left( y_2 - x_2 - \frac{2(y_1 - x_1 - hx_2)}{h} \right)^2 \right. \\
&\quad \left. + 5 \left( y_2 - x_2 - \frac{6(y_1 - x_1 - hx_2)}{h} + \frac{12(y_0 - x_0 - hx_1 - \frac{h^2x_2}{2})}{h^2} \right)^2 \right].
\end{aligned}$$

For  $n = 4$

$$\begin{aligned}
B &= \begin{pmatrix} 0 & 0 & 0 & -5040 \\ 0 & 0 & 720 & 5040h \\ 0 & -120 & -720h & -2520h^2 \\ 24 & 120h & 360h^2 & 840h^3 \end{pmatrix}, \\
A &= \begin{pmatrix} h^4 & h^5 & h^6 & h^7 \\ 4h^3 & 5h^4 & 6h^5 & 7h^6 \\ 12h^2 & 20h^3 & 30h^4 & 42h^5 \\ 24h & 60h^2 & 120h^3 & 210h^4 \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} \frac{35}{h^4} & -\frac{15}{h^3} & \frac{5}{2h^2} & -\frac{1}{6h} \\ -\frac{84}{h^5} & \frac{39}{h^4} & -\frac{7}{h^3} & \frac{1}{2h^2} \\ \frac{70}{h^6} & -\frac{34}{h^5} & \frac{13}{2h^4} & -\frac{1}{2h^3} \\ -\frac{20}{h^7} & \frac{10}{h^6} & -\frac{2}{h^5} & \frac{1}{6h^4} \end{pmatrix}, \\
L &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{4}{h} & 1 & 0 & 0 \\ \frac{12}{h^2} & \frac{8}{h} & 1 & 0 \\ \frac{24}{h^3} & \frac{36}{h^2} & \frac{12}{h} & 1 \end{pmatrix}, \quad U = \begin{pmatrix} h^4 & h^5 & h^6 & h^7 \\ 0 & h^4 & 2h^5 & 3h^6 \\ 0 & 0 & 2h^4 & 6h^5 \\ 0 & 0 & 0 & 6h^4 \end{pmatrix}, \\
L^{-1} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\frac{4}{h} & 1 & 0 & 0 \\ \frac{20}{h^2} & -\frac{8}{h} & 1 & 0 \\ -\frac{120}{h^3} & \frac{60}{h^2} & -\frac{12}{h} & 1 \end{pmatrix}, \quad U^{-1} = \begin{pmatrix} \frac{1}{h^4} & -\frac{1}{h^3} & \frac{1}{2h^2} & -\frac{1}{6h} \\ 0 & \frac{1}{h^4} & -\frac{1}{h^3} & \frac{1}{2h^2} \\ 0 & 0 & \frac{1}{2h^4} & -\frac{1}{2h^3} \\ 0 & 0 & 0 & \frac{1}{6h^4} \end{pmatrix},
\end{aligned}$$

and

$$\begin{aligned}
& C_{4,h}(x_0, x_1, x_2, x_3; y_0, y_1, y_2, y_3) \\
&= \frac{1}{h} \left[ (y_3 - x_3)^2 + 3 \left( y_3 - x_3 - \frac{2(y_2 - x_2 - hx_3)}{h} \right)^2 \right] \\
&+ \frac{5}{h} \left( y_3 - x_3 - \frac{6(y_2 - x_2 - hx_3)}{h} + \frac{12(y_1 - x_1 - hx_2 - \frac{h^2 x_3}{2})}{h^2} \right)^2 \\
&+ \frac{7}{h} \left( y_3 - x_3 - \frac{12(y_2 - x_2 - hx_3)}{h} + \frac{60(y_1 - x_1 - hx_2 - \frac{h^2 x_3}{2})}{h^2} \right. \\
&\quad \left. - \frac{120(y_0 - x_0 - hx_1 - \frac{h^2 x_2}{2} - \frac{h^3 x_3}{6})}{h^3} \right)^2.
\end{aligned}$$

We observe that these cost functions satisfy a property that

$$\begin{aligned}
C_{n,h}(x_0, x_1, \dots, x_{n-1}; y_0, y_1, \dots, y_{n-1}) &= C_{n-1,h}(x_1, \dots, x_{n-1}; y_1, \dots, y_{n-1}) \\
&+ \left| \sum_{j=0}^{n-1} \frac{\alpha_j}{h^{n-1-j}} \left[ y_j - \sum_{i=j}^{n-1} \frac{h^{i-j}}{(i-j)!} x_i \right] \right|^2,
\end{aligned}$$

for some  $\{\alpha_j\}_{j=0}^{n-1}$ , which is in accordance with Lemma 2.1.

## Acknowledgements

M. H. Duong was supported by ERC Starting Grant 335120. We would like to thank referees for their suggestions to improve the presentation of the paper.

## References

- [1] D. G. Aronson. Bounds for the fundamental solution of a parabolic equation. *Bulletin of the American Mathematical Society*, 73, 890–896, 1967.
- [2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics. ETH Zürich. Birkhauser, Basel, 2nd edition, 2008.
- [3] A. Bostan and P. Dumas. Wronskians and linear independence. *The American Mathematical Monthly*, 117(8):722–727, 2010.
- [4] J. C. Canuto. *Biomechanical online signature modeling applied to verification*. Theses, Institut National des Télécommunications, 2014.

- [5] J. Canuto, B. Dorizzi, and J. Montalvao. Dynamic signatures representation using the minimum jerk principle. In *Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*, Pages 1–6, 2013.
- [6] F. Cavalletti, M. Sedjro, and M. Westdickenberg. A variational time discretization for the compressible euler equations. *Arxiv: 1411.1012*, 2014.
- [7] M. H. Duong, A. Lamacz, M. A. Peletier, and U. Sharma. Variational approach to coarse-graining of generalized gradient flows. *Arxiv: 1507.03207*, 2015.
- [8] F. Delarue and S. Menozzi. Density estimates for a random noise propagating through a chain of differential equations. *Journal of Functional Analysis*, 259(6):1577–1630, 2010.
- [9] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic modelling and applied probability*. Springer, New York, NY, USA, 2nd edition, 1987.
- [10] M. H. Duong, M. A. Peletier, and J. Zimmer. GENERIC formalism of a Vlasov-Fokker-Planck equation and connection to large-deviation principles. *Nonlinearity*, 26(11):2951–2971, 2013.
- [11] M. H. Duong, M. A. Peletier, and J. Zimmer. Conservative-dissipative approximation schemes for a generalized kramers equation. *Mathematical Methods in the Applied Sciences*, 37(16):2517–2540, 2014.
- [12] M. H. Duong. Long time behaviour and particle approximation of a generalised Vlasov dynamic. *Nonlinear Analysis: : Theory, Methods & Applications*, 127:1–16, 2015.
- [13] M. H. Duong and M. H. Tran. On the fundamental solution and a variational formulation of a degenerate diffusion of Kolmogorov type. *In preparation*, 2017.
- [14] L. Desvillettes and C. Villani. On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems: The linear fokker-planck equation. *Communications on Pure and Applied Mathematics*, 54(1):1–42, 2001.
- [15] S. E. Engelbrecht. Minimum principles in motor control. *Journal of Mathematical Psychology*, 45:497–542, 2001.
- [16] T. Flash and N. Hogans. The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience*, 5:1688–1703, 1985.
- [17] P. Freeman. *Minimum Jerk Trajectory Planning for Trajectory Constrained Redundant Robots*. PhD thesis, Washington University in St. Louis, 2012.
- [18] W. Gangbo and M. Westdickenberg. Optimal transport for the system of isentropic Euler equations. *Communication in Partial Differential Equations*, 34(7-9):1041–1073, 2009.

- [19] N. Heckman. The theory and application of penalized methods or reproducing kernel hilbert spaces made easy. *Statistics Surveys*, 6:113–141, 2012.
- [20] N. Hogan. An organizing principle for a class of voluntary movements. *Journal of Neuroscience*, 4:2745–2754, 1984.
- [21] L. Hörmander. Hypoelliptic second order differential equations. *Acta Mathematica*, 119:147–171, 1967.
- [22] C. Huang. A variational principle for the Kramers equation with unbounded external forces. *Journal of Mathematical Analysis and Applications*, 250(1):333–367, 2000.
- [23] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the fokker-planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.
- [24] A. N. Kolmogorov. Zufällige Bewegungen. *Annals of Mathematics*, 35(2):116–117, 1934.
- [25] M. J. E. Richardson and T. Flash. Comparing smooth arm movements with the two-thirds power law and the related segmented-control hypothesis. *Journal of Neuroscience*, 22(18):8201–8211, 2002.
- [26] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
- [27] C. Villani. *Optimal transport : old and new*. Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2009.
- [28] M. Westdickenberg. Projections onto the cone of optimal transport maps and compressible fluid flows. *Journal of Hyperbolic Differential Equations*, 7(4):605–649, 2010.